BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 2: e25912
doi: 10.3897/biss.2.25912

OPEN ACCESS

Conference Abstract

Managing a Mass Digitization Project at Meise Botanic Garden: From Start to Finish

Henry Engledow‡, Sofie De Smedt‡, Quentin Groom‡, Ann Bogaerts‡, Piet Stoffelen‡, Marc Sosef‡, Paul Van Wambeke‡

‡ Meise Botanic Garden, Meise, Belgium

Corresponding author: Sofie De Smedt (sofie.desmedt@plantentuinmeise.be)

Received: 17 Apr 2018 | Published: 15 Jun 2018

Citation: Engledow H, De Smedt S, Groom Q, Bogaerts A, Stoffelen P, Sosef M, Van Wambeke P (2018) Managing a Mass Digitization Project at Meise Botanic Garden: From Start to Finish. Biodiversity Information Science and Standards 2: e25912. https://doi.org/10.3897/biss.2.25912

Abstract

Mass digitization is a large undertaking for a collection. It is disruptive of routine and can challenge long-held practises. Having been through the procedure and survived, we feel we have a lot of experience to share with other institutions who are considering taking on this challenge. The changes that digitization has made to our institution are positive and the digitization a success, but that is not to say that we would not have done some things differently, were we to repeat the exercise.

In 2015 Meise Botanic Garden received a grant from the Flemish Government to upgrade its digitization infrastructure and mass digitize 1.2 million specimens from its African and Belgian Herbaria. The new infrastructure improved our workflow significantly, enabling us to digitize specimens five to ten times faster while also improving their quality.

The mass digitization part of the project was split into two parts, imaging and transcription. The contract was awarded and out-sourced to Picturae, who started imaging in May 2016 using a conveyor belt installation. Prior to starting, a significant amount of preparation was required at the herbarium. Within one year, 1.2 million specimens were imaged. The images were captured as TIFF files and stored in triplicate at The Flemish Institute for Archiving (VIAA), while smaller derived JPEG 2000 and JPEG files were generated for day-to-day use.

© Engledow H et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The second part of the project was label transcription. A third of the specimens were transcribed in-house for capturing minimal data (barcode, filing name, collector, collector number & country of origin). This was partly done to reduce costs, but also allowed us to compare in-house to out-sourced transcription. Some 500,000 specimens were transcribed, either completely or partially, by Alembo (subcontracted by Picturae).The remaining 200.000 specimens from our Belgian Herbarium are being transcribed using crowdsourcing. The latter is being realized through the citizen science platform *DoeDat* (www.doedat.be) that was launched in November 2017.

Many lessons have been learnt with respect to implementing mass digitization, both practically and sociologically. Many of the problems encountered during the project could have been avoided by changing the workflow. The addition of extra control points during the process could have reduced problems encountered later in the data capture process. Solving these problems at a later stage was time consuming. Trying to “save money” can result in a disruptive workflow, which may lead to a number of costly errors. Mass digitization has fundamentally changed the workflow in our collections and the way in which our herbarium is managed. All images for the African and Belgian collections may be now found on our new virtual herbarium www.botanicalcollections.be.

Keywords

Mass Digitization, transcription, imaging, crowdsourcing, herbarium

Presenting author

Henry Engledow

Presented at

SPNHC 2018